Asymmetric Laplace Regression: Maximum Likelihood, Maximum Entropy and Quantile Regression

Article · January 2015

DOI: 10.1515/jem-2014-0018

4 authors, including:

Anil K. Bera
University of Illinois, Urbana-Champaign
120 PUBLICATIONS **10,590** CITATIONS

SEE PROFILE

Gabriel Montes-Rojas
National Scientific and Technical Research Council
86 PUBLICATIONS **953** CITATIONS

SEE PROFILE

Sung Y. Park
Chung-Ang University
40 PUBLICATIONS **559** CITATIONS

SEE PROFILE

Research Article

Anil K. Bera, Antonio F. Galvao Jr., Gabriel V. Montes-Rojas* and Sung Y. Park

Asymmetric Laplace Regression: Maximum Likelihood, Maximum Entropy and Quantile Regression

Abstract: This paper studies the connections among the asymmetric Laplace probability density (ALPD), maximum likelihood, maximum entropy and quantile regression. We show that the maximum likelihood problem is equivalent to the solution of a maximum entropy problem where we impose moment constraints given by the joint consideration of the mean and median. The ALPD score functions lead to joint estimating equations that delivers estimates for the slope parameters together with a representative quantile. Asymptotic properties of the estimator are derived under the framework of the quasi maximum likelihood estimation. With a limited simulation experiment we evaluate the finite sample properties of our estimator. Finally, we illustrate the use of the estimator with an application to the US wage data to evaluate the effect of training on wages.

Keywords: asymmetric Laplace distribution; quantile regression; treatment effects.

JEL Classification: C14; C31.

DOI 10.1515/jem-2014-0018
Previously published online March 3, 2015

1 Introduction

Different choices of loss functions determine different ways of defining the location of a random variable Y. For example, squared, absolute value, and step function lead to mean, median and mode, respectively (see Manski 1991, for a general discussion). For a given quantile $\tau \in (0, 1)$, consider the loss function in a standard quantile estimation problem,

$$L_{1,n}(\mu;\tau)=\sum_{i=1}^{n}\rho_\tau(y_i-\mu)=\sum_{i=1}^{n}(y_i-\mu)(\tau-1(y_i\leq\mu)), \tag{1}$$

as proposed by Koenker and Bassett (1978). Minimizing $L_{1,n}$ with respect to the location parameter μ is identical to maximizing the likelihood based on the asymmetric Laplace probability density (ALPD):

$$f(y;\mu,\tau,\sigma)=\frac{\tau(1-\tau)}{\sigma}\exp\left(-\frac{\rho_\tau(y-\mu)}{\sigma}\right), \tag{2}$$

***Corresponding author: Gabriel V. Montes-Rojas,** Department of Economics, City University London, 10 Northampton Square, London EC1V 0HB, UK, E-mail: Gabriel.Montes-Rojas.1@city.ac.uk
Anil K. Bera: Department of Economics, University of Illinois, 1407 W. Gregory Drive, Urbana, IL 61801, USA
Antonio F. Galvao Jr.: Department of Economics, University of Iowa, W334 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242, USA
Sung Y. Park: School of Economics, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul, Korea

for given (τ, σ). The well known symmetric Laplace (double exponential) distribution is a special case of (2) when $\tau=1/2$.

Interestingly, the parameter μ in functions (1) and (2) is at the same time the location parameter, the τ-th quantile, and the mode of the ALPD. For the simple (unconditional) case, the minimization of (1) for a given τ returns the corresponding order-statistic. For example, if we set $\tau=\{0.1, 0.2,\ldots,0.9\}$, the solutions are, respectively, the nine deciles of Y. Using the ALPD, for a given τ, maximization of the corresponding likelihood function also gives that particular order statistic. The main idea of this paper is to *jointly* estimate τ and the corresponding order statistic of Y which can be taken as a good summary statistic of the data. The ALPD case naturally provides a framework for this. If the true distribution is not ALPD (being ALPD is not a requirement for the purposes of this paper), the selected τ-quantile does not necessarily lead to the mode, but to a point estimate that maximizes the likelihood function. The above notion is extended to modeling the "conditional location" of Y given covariates X.

This provides a new interpretation of quantile regression (QR). In a related context, Zou and Yuan (2008) develop the composite quantile regression (CQR) estimator where the slope parameters of interest are obtained from a mixture of the objective functions from different quantile regression models. They show that by combining across multiple quantile regression models, improvements in terms of estimation efficiency are obtained. Zhao and Xiao (2011) argue that it is crucial to optimally combine information over quantiles for efficient estimation. The ALPD framework is a particular case of CQR if we consider an objective function based on a continuous of quantiles, $\tau\in(0, 1)$, which are weighted by a function of τ, $\tau(1-\tau)$. By estimating τ through maximization of the likelihood function, the ALPD model provides a twist to the QR problem, as now τ becomes the *most likely quantile* in a regression set-up. Moreover, we show that the ALPD framework can be seen as a penalized quantile optimization function, where the penalty is also based on $\tau(1-\tau)$. The penalty can thus be interpreted as the cost of deviating from the median.

We show that the score functions implied by the ALPD maximum likelihood (ML) estimation are not restricted to the true data generating process being ALPD, but they arise as the solution to a maximum entropy (ME) problem where we impose moment constraints given by the joint consideration of the mean and median. By so doing, the ALPD-ML estimator combines the information in the mean and the median to capture the asymmetry of the underlying empirical distribution (see Park and Bera 2009, for a discussion). We propose a Z-estimator that is based on the estimating equations from the ML score functions (which also correspond to the ME problem). We refer to this estimator as ZQR. We derive the asymptotic properties of the estimator by showing consistency and asymptotic normality under certain regularity conditions.

The Z-estimator does not impose that the distribution is ALPD. Thus, although the original motivation for using the estimating equations is based on the ALPD, the final estimator is independent of this requirement. Moreover, the Z-estimator is studied for independent but not identically distributed observations, where QR provides a very useful way to study heterogeneity in the data.

The parameter interpretation is also independent of the ALPD requirement. The proposed estimator has an interesting interpretation from a policy perspective. The QR analysis gives a full range of estimators that account for heterogeneity in the response variable to certain covariates. The proposed estimator answers the question: of all the heterogeneity in the conditional regression model, which one can be taken as representative? In general, the entire QR process is of interest because we would like to either test global hypotheses about conditional distributions or make comparisons across different quantiles (for a discussion about inference in QR models see Koenker and Xiao 2002). However, selecting a particular quantile provides an estimator as parsimonious as the ordinary least squares (OLS) or median estimators. Note that the proposed estimator should not be viewed as a substitute of the QR analysis, but only as a complement.

Several studies developed the properties of the ML estimators based on ALPD. Hinkley and Revankar (1997) derived the asymptotic properties of the unconditional MLE. Kotz, Kozubowski, and Podgórsk (2002b) and Yu and Zhang (2005) considered alternative MLE approaches for ALPD. Moreover, models based on ALPD have been proposed in different contexts. For instance, Machado (1993) used the ALPD to derive a Schwartz information criterion for model selection for QR models, and Koenker and Machado (1999) introduced a goodness-of-fit measure for QR and related inference processes. Yu and Moyeed (2001) and Geraci

and Botai (2007) used a Bayesian QR approach based on the ALPD. Komunjer (2005) constructed a new class of estimators for conditional quantiles in possibly misspecified nonlinear models with time series data. The estimators belong to the family of quasi-maximum likelihood estimators (QMLEs) and are based on a family of 'tick-exponential' densities. Under the asymmetric Laplace density, the corresponding QMLE reduces to the Koenker and Bassett (1978) linear quantile regression estimator. In addition, Komunjer (2007) developed a parametric estimator for the risk of financial time series expected shortfall based on the asymmetric power distribution, derived the asymptotic distribution of the MLE, and constructed a consistent estimator for its asymptotic covariance matrix.

Finally, we illustrate the implementation of the proposed ZQR estimator. We apply the estimator to the estimation of quantile treatment effects of subsidized training on wages under the Job Training Partnership Act (JTPA). We discuss the relationship between OLS, median regression and ZQR estimates of the JTPA treatment effect. We show that each estimator provides different treatment effect estimates. Moreover, we extend our ZQR estimator to Chernozhukov and Hansen (2006, 2008) instrumental variables strategy in QR.

The rest of the paper is organized as follows. Section 2 develops the ML and ME frameworks of the problem. Section 3 considers a Z-estimator of QR and derives its the asymptotic distribution. In Section 4 we report a small Monte Carlo study to assess the finite sample performance of the estimator. Section 5 deals with an empirical illustration to the effect of training on wages. Finally, conclusions are in the last section.

2 Maximum Likelihood and Maximum Entropy

In this section we describe the ML problem based on the ALPD and show its connection with the ME. In the next section we will propose a Z-estimator based on the resulting estimating equations from the ML-ME problem.

2.1 Maximum Likelihood

Using (2), consider the maximization of the log-likelihood function of an ALPD:

$$L_{2,n}(\mu,\tau,\sigma)=n\ln\left(\frac{1}{\sigma}\tau(1-\tau)\right)-\sum_{i=1}^{n}\frac{1}{\sigma}\rho_\tau(y_i-\mu)=n\ln\left(\frac{1}{\sigma}\tau(1-\tau)\right)-\frac{1}{\sigma}L_{1,n}(\mu;\tau), \tag{3}$$

with respect to μ, τ and σ. The first order conditions from (3) lead to the following estimating equations (EE):

$$\sum_{i=1}^{n}\frac{1}{\sigma}\left(\frac{1}{2}\text{sign}(y_i-\mu)+\tau-\frac{1}{2}\right)=0, \tag{4}$$

$$\sum_{i=1}^{n}\left(\frac{1-2\tau}{\tau(1-\tau)}-\frac{(y_i-\mu)}{\sigma}\right)=0, \tag{5}$$

$$\sum_{i=1}^{n}\left(-\frac{1}{\sigma}+\frac{1}{\sigma^2}\rho_\tau(y_i-\mu)\right)=0. \tag{6}$$

The structure of the estimating functions suggests that the solution to the ML problem can be obtained by first obtaining every quantile, say $\hat{\mu}(\tau)$, then constructing $\hat{\sigma}(\tau)=\frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i-\hat{\mu}(\tau))$, and then plugging both $\hat{\mu}(\tau)$ and $\hat{\sigma}(\tau)$ in (5) to obtain $\hat{\tau}$. Thus the solution is $(\hat{\mu},\hat{\tau},\hat{\sigma})=(\hat{\mu}(\hat{\tau}),\hat{\tau},\hat{\sigma}(\hat{\tau}))$.

In the Appendix A we interpret the parameters (μ,τ,σ) as the solution to this system of equations (4)–(6), which corresponds to the Z-estimator proposed in Section 3. This interpretation does not require the true data generating process to be i.i.d. ALPD. In general, σ provides a measure of dispersion given by the expectation

of the QR check function. That is, similar to the least squares case, σ can be interpreted as the expected value of the loss function. Note that in the independent but non identically distributed case σ does not necessarily correspond to a measure of the dispersion of the error term, which may not be the same across individuals. Rather it provides an inverse measure of precision of the parameter estimates. τ represents a measure of skewness of the data that combines the information in the mean and the median to capture the underlying asymmetry of the error distribution, and μ is the associated τ-quantile. In the i.i.d. ALPD case μ is also the mode of Y, and by selecting the mode of the distribution, the problem solves for the order statistic that has minimum variance. The reason is that $Var(\hat{\mu}(\tau)) \propto \frac{\tau(1-\tau)}{f^2(\mu(\tau))}$. A more general interpretation of μ is that it is the location parameter that is most probable, in the sense it maximizes the entropy. We develop this idea in the next subsection.

Interestingly, the loss function corresponding to (3) can be rewritten as a two-parameter loss function

$$-\frac{1}{n}L_{2,n}(\mu,\tau)=\ln\left(\frac{1}{n}L_{1,n}(\mu;\tau)\right)-\ln(\tau(1-\tau)). \tag{7}$$

This determines that $L_{2,n}(\mu, \tau, \sigma)$ can be seen as a penalized quantile optimization function, where we minimize $\ln\left(\frac{1}{n}L_{1,n}(\mu;\tau)\right)$ and penalize it by $-\ln(\tau(1-\tau))$. The penalty can be interpreted as the cost of deviating from the median, i.e., for $\tau=1/2$, $-\ln(\tau(1-\tau))=-\ln(1/4)$ is the minimum, while for either $\tau\to 0$ or $\tau\to 1$ the penalty goes to $+\infty$.

2.2 Maximum Entropy

The ALPD can be characterized as a maximum entropy density obtained by maximizing Shannon's entropy measure subject to two moment constraints (see Kotz, Kozubowski, and Podgórsk 2002a):

$$f_{ME}(y)\equiv \arg\max_f -\int f(y)\ln f(y)dy \tag{8}$$

subject to

$$E|y-\mu|=c_1, \tag{9}$$

$$E(y-\mu)=c_2, \tag{10}$$

and the normalization constraint, $\int f(y)dy=1$, where c_1 and c_2 are known constants. The solution to the above optimization problem using the Lagrangian has the familiar exponential form

$$f_{ME}(y:\mu,\lambda_1,\lambda_2)=\frac{1}{\Omega(\theta)}\exp[-\lambda_1|y-\mu|-\lambda_2(y-\mu)],\ -\infty<y<\infty, \tag{11}$$

where λ_1 and λ_2 are the Lagrange multipliers corresponding to the constraints (9) and (10), respectively, $\theta=(\mu, \lambda_1, \lambda_2)'$ and $\Omega(\theta)$ is the normalizing constant. Note that $\lambda_1\in\mathbb{R}^+$ and $\lambda_2\in[-\lambda_1, \lambda_1]$ so that $f_{ME}(y)$ is well-defined. Symmetric Laplace density (SLD) is a special case of ALPD when λ_2 is equal to zero.

Interestingly, the constraints (9) and (10) capture, respectively, the dispersion and asymmetry of the ALPD. The marginal contribution of (10) is measured by the Lagrangian multiplier λ_2. If λ_2 is close to 0, then (10) does not have useful information for the data, and therefore, SLD is the most appropriate. In this case, μ is known to be the median of the distribution. On the other hand, when λ_2 is not zero, it measures the degree of asymmetry of the ME distribution. Thus the non-zero value of λ_2 makes $f_{ME}(\cdot)$ to deviate from the SLD, and therefore, changes the location, μ, of the distribution to adhere the maximum value of the entropy (for general notion of entropy, see Soofi and Retzer 2002).

Let us write (9) and (10), respectively, as

$$\int \phi_1(y,\mu) f_{ME}(y:\mu,\lambda_1,\lambda_2)dy=0 \text{ and } \int \phi_2(y,\mu) f_{ME}(y:\mu,\lambda_1,\lambda_2)dy=0,$$

where $\phi_1(y,\mu)=|y-\mu|-c_1$ and $\phi_2(y,\mu)=(y-\mu)-c_2$. By substituting the solution $f_{ME}(y:\mu,\lambda_1,\lambda_2)$ into the Lagrangian of the maximization problem in (8), we obtain the profiled objective function

$$h(\lambda_1,\lambda_2,\mu)=\ln\int \exp\left[-\sum_{j=1}^{2}\lambda_j\phi_j(y,\mu)\right]dy. \tag{12}$$

The parameters λ_1, λ_2 and μ can be estimated by solving the following saddle point problem (Kitamura and Stutzer 1997)

$$\hat{\mu}_{ME}=\arg\max_{\mu}\ln\int\exp\left[-\sum_{j=1}^{2}\hat{\lambda}_{j,ME}\phi_j(y,\mu)\right]dy,$$

where $\hat{\lambda}_{ME}=(\hat{\lambda}_{1,ME},\hat{\lambda}_{2,ME})$ is given by

$$\hat{\lambda}_{ME}(\mu)=\arg\min_{\lambda}\ln\int\exp\left[-\sum_{j=1}^{2}\lambda_j\phi_j(y,\mu)\right]dy.$$

Solving the above saddle point problem is relatively easy since the profiled objective function has the exponential form. However, generally, c_1 and c_2 are not known functions of parameters and Lagrange multipliers in a non-linear fashion. Moreover, in some cases, the closed form of c_1 and c_2 is not known. In order to deal with this problem, we simply consider the sample counterparts, say, $c_1=(1/n)\sum_{i=1}^{n}|y_i-\mu|$ and $c_2=(1/n)\sum_{i=1}^{n}(y_i-\mu)$. Then, it can be easily shown that the profiled objective function is simply the negative log-likelihood function of asymmetric Laplace density, i.e., $h(\lambda_1,\lambda_2,\mu)=-(1/n)L_{2,n}(\mu,\tau,\sigma)$ (see Ebrahimi, Soofi, and Soyer 2008). In this case, $\hat{\mu}_{ME}$ and $\hat{\lambda}_{ME}$ satisfy the following first order conditions $\partial h/\partial\mu=0$, $\partial h/\partial\lambda_2=0$ and $\partial h/\partial\lambda_1=0$, respectively:

$$-\frac{\lambda_1}{n}\sum_{i=1}^{n}\text{sign}(y_i-\mu)-\lambda_2=0. \tag{13}$$

$$\frac{2\lambda_2}{(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)}+\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)=0, \tag{14}$$

$$-\frac{1}{\lambda_1}\frac{\lambda_1^2+\lambda_2^2}{\lambda_1^2-\lambda_2^2}+\frac{1}{n}\sum_{i=1}^{n}|y_i-\mu|=0. \tag{15}$$

Equations (13)–(15) are a re-parameterized version of (4)–(6). In fact, from a comparison of (2) and (11) we can see that $\lambda_1=1/(2\sigma)$, $\lambda_2=(2\tau-1)/(2\sigma)$ and $\Omega(\theta)=\sigma/(\tau(1-\tau))$. Given λ_1, the degree of asymmetry is explained by λ_2 that is proportionally equal to $2\tau-1$ in ALPD. Note that $\lambda_2=0$ when $\tau=0.5$, i.e., μ is the median. Thus finding the most appropriate degree of asymmetry is equivalent to estimating τ based on the ML method.

The role of the two moment constraints can be explained by the linear combination of two moment functions, $|y-\mu|$ and $(y-\mu)$. Figure 1 plots $g(y;\lambda_1,\lambda_2,\mu)=\lambda_1|y-\mu|+\lambda_2(y-\mu)$ with three different values of λ_2, $\lambda_1=1$, and $\mu=0$. In general, $g(y;\lambda_1,\lambda_2,\mu)$ can be seen as a loss function. Clearly, this loss function is symmetric when $\lambda_2=0$. When $\lambda_2=1/3$, $g(\cdot)$ is tilted so that it puts more weight on the positive values in order to attain the maximum of the Shannon's entropy (and the reverse is true for $\lambda_2=-1/3$).

3 A New Z-Estimator

Now we consider the conditional version of the model describe previously, as a linear model $y=x'\beta+u$, where x refers to a p-vector of exogenous covariates, the parameter of interest is $\beta\in\mathbb{R}^p$, and u denotes the unobserv-

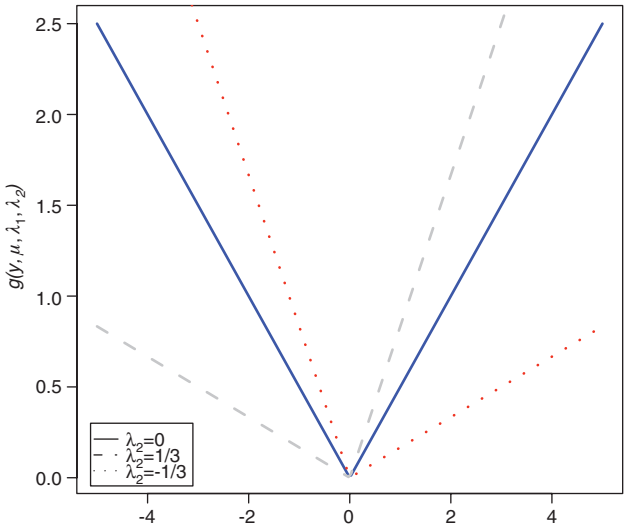

Figure 1: Linear combination of $|y-\mu|$ and $(y-\mu)$.

able component. As noted in Angrist, Chernozhukov, and Fernández-Val (2006), QR provides the best linear predictor for y under the asymmetric loss function

$$L_{3,n}(\beta;\tau)=\sum_{i=1}^{n}\rho_\tau(y_i-x_i'\beta)=\sum_{i=1}^{n}((y_i-x_i'\beta)(\tau-1(y_i\le x_i'\beta))), \tag{16}$$

where β is assumed to be a function of the *fixed* quantile τ of the unobservable components, that is $\beta(\tau)$. If u is assumed to follow an ALPD, the log-likelihood function is

$$\begin{aligned}L_{4,n}(\beta,\tau,\sigma)&=n\ln\left(\frac{1}{\sigma}\tau(1-\tau)\right)-\sum_{i=1}^{n}\left(\frac{1}{\sigma}\rho_\tau(y_i-x_i'\beta)\right)\\&=n\ln\left(\frac{1}{\sigma}\tau(1-\tau)\right)-\frac{1}{\sigma}L_{3,n}(\beta;\tau).\end{aligned} \tag{17}$$

Estimating β in this framework provides the marginal effect of x on the conditional τ-quantile of y.

We have interest in estimating $\theta=(\beta,\tau,\sigma)'$, and that can be obtained by solving the EE, $\nabla L_{4,n}(\beta,\tau,\sigma)=0$, i.e.,

$$\frac{\partial L_{4,n}(\beta,\tau,\sigma)}{\partial\beta}=\sum_{i=1}^{n}\frac{1}{\sigma}\left(\frac{1}{2}\text{sign}(y_i-x_i'\beta)+\tau-\frac{1}{2}\right)x_i=0, \tag{18}$$

$$\frac{\partial L_{4,n}(\beta,\tau,\sigma)}{\partial\tau}=\sum_{i=1}^{n}\left(\frac{1-2\tau}{\tau(1-\tau)}-\frac{(y_i-x_i'\beta)}{\sigma}\right)=0, \tag{19}$$

$$\frac{\partial L_{4,n}(\beta,\tau,\sigma)}{\partial\sigma}=\sum_{i=1}^{n}\left(-\frac{1}{\sigma}+\frac{1}{\sigma^2}\rho_\tau(y_i-x_i'\beta)\right)=0, \tag{20}$$

which are the counterparts of (4)–(6), the EE for the location-scale model. The implementation of the estimator in practice is simple. An iteration algorithm can be used to solve for the estimates in the above estimating equations. Computationally, the estimates can be obtained by constructing a grid for quantiles τ and solving the QR problem as in (18) and (20) to find $\hat{\beta}(\tau)$ and $\hat{\sigma}(\tau)$ for each τ. Finally, obtain $\hat{\tau}$ that finds an approximate zero in (19), and from this step recover the estimates of $\hat{\beta}(\tau)$ and $\hat{\sigma}(\tau)$ as $\hat{\beta}(\hat{\tau})$ and $\hat{\sigma}(\hat{\tau})$, respectively. This algorithm is similar to the one proposed in Hinkley and Revankar (1997) and Yu and Zhang (2005) that

compute the estimators for ML under the ALPD. In our implementation, we find that the algorithm converges fast and is very precise.

As we stated before, $L_{4,n}$ can be written as a penalized QR problem loss function that depends only on (β, τ):

$$-\frac{1}{n}L_{4,n}(\beta,\tau)=\ln\left(\frac{1}{n}L_{3,n}(\beta;\tau)\right)-\ln(\tau(1-\tau)), \tag{21}$$

and the interpretation is the same as discussed in Section 2.1.

We propose a Z-estimator based on the score functions from equations (18)–(20). Thus, although the original motivation for using the estimating equations is based on the ALPD, the final estimator is independent of that requirement. Now we define the estimating functions

$$\psi_\theta(y,x)=\begin{pmatrix}\psi_{1\theta}(y,x)\\ \psi_{2\theta}(y,x)\\ \psi_{3\theta}(y,x)\end{pmatrix}=\begin{pmatrix}\frac{1}{\sigma}(\tau-1(y<x'\beta))x\\ \frac{1-2\tau}{\tau(1-\tau)}-\frac{(y-x'\beta)}{\sigma}\\ -\frac{1}{\sigma}+\frac{1}{\sigma^2}\rho_\tau(y-x'\beta)\end{pmatrix},$$

and the corresponding EE

$$\Psi_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\frac{1}{\sigma}(\tau-1(y_i<x_i'\beta))x_i\\ \frac{1-2\tau}{\tau(1-\tau)}-\frac{(y_i-x_i'\beta)}{\sigma}\\ -\frac{1}{\sigma}+\frac{1}{\sigma^2}\rho_\tau(y_i-x_i'\beta)\end{pmatrix}=\frac{1}{n}\sum_{i=1}^{n}\psi_\theta(y_i,x_i)=0.$$

The Z-estimator $\hat{\theta}_n$ is the approximate zero of the above data-dependent function that satisfies $||\Psi_n(\hat{\theta}_n)||\xrightarrow{p}0$, where $||\cdot||$ is the Euclidean norm.

The interpretation of the parameters is similar to that given in Section 2.1 and Appendix A. In the proposed Z-estimator the interpretation of the parameter β is analogous to the interpretation of the location parameter in the QR literature. As in the least squares case, the scale parameter σ can be interpreted as the expected value of the loss function, which in the QR case corresponds to the expectation of the $\rho_\tau(\cdot)$ function. Finally, τ captures the asymmetry of the conditional distribution of $y|x$ and is associated with the "most probable" quantile, in the sense it maximizes the entropy. Thus, this approach delivers estimates for the slope parameters together with the associated "most probable" quantile. As in the simple unconditional case in Section 2.1 for the ALPD-ML, the proposed estimator combines the information in the mean and the median to capture the asymmetry of the underlying innovations distribution.

We introduce the following assumptions to derive the asymptotic properties.

Assumption A1 Let $y_i=x_i'\beta_0+u_i$, where (y_i, x_i) is independent across i, $i=1, 2,\ldots,n$.

Assumption A2 The conditional distribution function of y_i, $G_i(y|x)$, is absolutely continuous with conditional densities, $g_i(y|x)$, with $0<g_i(\cdot|\cdot)<\infty$.

Assumption A3 Let Θ be a compact set, with $\theta=(\beta,\tau,\sigma)'\in\Theta$, where $\beta\in\mathcal{B}\subset\mathbb{R}^p$, $\tau\in\mathcal{T}\subset(0,1)$, and $\sigma\in\mathcal{S}\subset\mathbb{R}^+$, and θ_0 is an interior point of Θ.

Assumption A4 $sup_i E||x_i||^{2+\epsilon}<\infty$, and $sup_i E||y_i||^{2+\epsilon}<\infty$ for some $\epsilon>0$.

Assumption A5 (i) Define $\Psi(\theta)=E[\psi_\theta(y, x)]$. Assume that $\Psi(\theta_0)=0$ for a unique $\theta_0\in\Theta$. (ii) Define $\Psi_n(\theta)=\mathbb{E}_n[\psi_\theta(y,x)]=\frac{1}{n}\sum_{i=1}^{n}\psi_\theta(y_i,x_i)$. Assume that $||\Psi_n(\hat{\theta}_n)||=o_p(n^{-1/2})$.

Assumption A1 is common in the QR literature, where the observations are assumed to be independent across individuals, but not necessarily identically distributed. This condition allows for heterogeneous effects [i.e., $\beta(\tau)$ varies across τ] in the quantile process. Assumption A2 restricts the conditional distribution of the dependent variable. Assumption A3 is standard in asymptotic theory and imposes compactness of the parameter space. A4 is needed to guarantee the asymptotic behavior of the estimator and is sufficient to encompass QR and least squares estimators asymptotic results in a non i.i.d. context. Finally, Assumption A5 imposes an identifiability condition and ensure that the solution to the estimating equations is "nearly-zero," and it deserves further discussion.

The first part of A5 imposes a unique solution condition. Similar restrictions are usually used in the QR literature to satisfy $E[\psi_{1\theta}(y, x)]=0$ for a unique β for any given τ. This condition also appears frequently in the M- and Z-estimators literatures. Uniqueness in QR is very delicate and is actually imposed. For instance (Chernozhukov, Fernández-Val, and Melly 2009, 49) propose an approximate Z-estimator for the QR process and assume that the true parameter $\beta_0(\tau)$ solves $E[(\tau-1\{y\leq X'\beta_0(\tau)\})X]=0$. Angrist, Chernozhukov, and Fernández-Val (2006) impose an assumption of the form: $\beta(\tau)=\arg\min_\beta E[\rho_\tau(y-x'\beta)]$ is unique (see for instance their Theorems 1 and 2). See also He and Shao (2000) and Schennach (2008) for related discussion.[1] Combining these QR assumptions and bounded moments we guarantee uniqueness for $E[\psi_{3\theta}(y, x)]=0$, because $\sigma(\tau)=E[\rho_\tau(y-x\beta(\tau))]$, σ is unique for any given τ.

The second equation $E[\psi_{2\theta}(y, x)]=0$ is satisfied if $\frac{1-2\tau}{\tau(1-\tau)}=\frac{E[y-x'\beta(\tau)]}{\sigma(\tau)}$. The uniqueness assumption implies that only one $\tau=\tau_0$ satisfies this condition. This condition however depends on the joint distribution of $u_i\equiv y_i-x_i'\beta_0$, $i=1,2,..,n$.[2] To clarify consider the following example. If $u\sim i.i.d.U(-1, 1)$, we do not expect to find a representative quantile, and in fact, $E[\psi_{2\theta}(y, x)]=0$ has more than one solution. However, as noted by an anonymous referee, typical unimodal distributions (i.e., of course ALPD, but also Gaussian, χ^2) deliver one unique solution, and in practice, most error distributions turn out to be unimodal. Multimodal distributions may also posses a unique solution which can be checked in practice by plotting $\frac{1}{n}\sum_{i=1}^n\psi_{2(\hat\beta(\tau),\tau,\hat\sigma(\tau))}(y_i, x_i)$ for all τ, where $\hat\beta(\tau)$ is the τ-QR estimator and $\hat\sigma(\tau)=\frac{1}{n}\sum_{i=1}^n\rho_\tau(y_i-x_i\hat\beta(\tau))$ and evaluating if it has one or more zeros. In the empirical application this is carried out in Figure 3.

The second part of A5 is used to ensure that the solution to the approximated working estimating equations is close to zero. The solution for the estimating equations, $\Psi_n(\hat\theta_n)=0$, does not hold in general. In most cases, this condition is actually equal to zero, but least absolute deviation of linear regression is one important exception. The indicator function in the first estimating equations determines that it may not have an exact zero. It is common in the literature to work with M- and Z-estimators $\hat\theta_n$ of θ_0 that satisfy $\sum_{i=1}^n\psi(x_i, \hat\theta_n)=o_p(\delta_n)$,

---

1 It is possible to impose more primitive conditions to ensure uniqueness. These conditions are explored and discussed in Theorem 2.1 in Koenker (2005, 36). If the y's have a bounded density with respect to Lebesgue measure then the observations (y, x) will be in general position with probability one and a solution exists. See definition 2.1 in Koenker (2005) for a definition of general position. However, uniqueness cannot be ensured if the covariates are discrete (e.g., dummy variables). If the x's have a component that have a density with respect to a Lebesgue measure, then multiple optimal solutions occur with probability zero and the solution is unique. However, these conditions are not very attractive, and uniqueness is in general imposed as an assumption.

2 Define $m(\tau)=\frac{1-2\tau}{\tau(1-\tau)}$ and $d(\tau)=\frac{E[y-x'\beta(\tau)]}{\sigma(\tau)}$. Note that $m(\tau)$ is a continuous strictly decreasing function, has a unique zero at $\tau=1/2$ and $m(\tau)>0$ for $\tau<1/2$, $m(\tau)<0$ for $\tau>1/2$. As $\tau\to0$, $m(\tau)\to+\infty$, and as $\tau\to1$, $m(\tau)\to-\infty$. Finally, $\frac{dm(\tau)}{d\tau}=\frac{-2\tau(1-\tau)-(1-2\tau)^2}{\tau^2(1-\tau)^2}=\frac{-2\tau+2\tau^2-1+4\tau-4\tau^2}{\tau^2(1-\tau)^2}=\frac{-1+2\tau-2\tau^2}{\tau^2(1-\tau)^2}=\frac{-1+2\tau(1-\tau)}{\tau^2(1-\tau)^2}<0$ for all $\tau\in(0, 1)$. $d(\tau)$ is also a continuous strictly decreasing function (because of Assumption A.2) with a unique zero at τ^* such that $E[y-x'\beta(\tau^*)]=0$, $d(\tau)>0$ for $\tau<\tau^*$ and $d(\tau)<0$ for $\tau>\tau^*$.

for some sequence δ_n. For example, Huber (1967) considered $\delta_n=\sqrt{n}$ for asymptotic normality, and Hinkley and Revankar (1997) verified the condition for the unconditional asymmetric double exponential case. This condition also appears in the QR literature, see for instance He and Shao (1996) and Wei and Carroll (2009). In the approximate Z-estimator for quantile process, Chernozhukov, Fernández-Val, and Melly (2009) had the empirical moment functions $\hat{\Psi}(\theta, u)=E_n[g(W_i, \theta, u)]$, for each $u\in T$, where the estimator $\hat{\theta}(u)$ satisfies $||\hat{\Psi}(\hat{\theta}(u), u)|| \leq \inf_{\theta\in\Theta}||\hat{\Psi}(\theta, u)||+\epsilon_n$ where $\epsilon_n=o(n^{-1/2})$. For the QR case, Koenker (2005, 36) commented that the absence of a zero to the problem $\Psi_{1n}(\hat{\beta}_n(\tau))=0$, where $\hat{\beta}_n(\tau)$ is the quantile regression optimal solution for a given τ and σ, "is unusual, unless the y_i's are discrete." Here we follow the standard conditions for M- and Z-estimators and impose A5(ii). For a more general discussion about this condition on M- and Z-estimators see for instance Kosorok (2008, 399–407).

Now we state the consistency property of the estimator.

Theorem 1 *Under Assumptions A1–A5,* $\left\|\hat{\theta}_n-\theta_0\right\|\xrightarrow{p}0$.

Proof: Define $\mathcal{F}\equiv\{\psi_\theta(y, x), \theta\in\Theta\}$, and recall that $\Psi_n(\theta)=\frac{1}{n}\sum_{i=1}^n\psi_\theta(y_i, x_i)$ and $\Psi(\theta)=E[\psi_\theta(y, x)]$. First note that, under conditions A3 and A5, the function $\Psi(\theta)$ satisfies,

$$\inf_{\theta:d(\theta,\theta_0)\geq\epsilon}||\Psi(\theta)||>0=||\Psi(\theta_0)||,$$

because for a compact set Θ and a continuous function Ψ, uniqueness of θ_0 as a zero implies this condition.

Now we need to show that $\sup_{\theta\in\Theta}||\Psi_n(\theta)-\Psi(\theta)||\xrightarrow{p}0$. By Lemma A1 in the Appendix B, we know that the class $\mathcal{F}$ is Donsker. Donskerness implies a uniform law of large numbers such that

$$\sup_{\theta\in\Theta}|\mathbb{E}_n[\psi_\theta(y, x)]-E[\psi_\theta(y, x)]|\xrightarrow{p}0,$$

where $f\mapsto\mathbb{E}_n[f(w)]=\frac{1}{n}\sum_{i=1}^n f(w_i)$. Hence we have $\sup_{\theta\in\Theta}||\Psi_n(\theta)-\Psi(\theta)||\xrightarrow{p}0$.

Finally, from assumptions A1–A5 the problem has a unique root and also we have $||\Psi_n(\hat{\theta}_n)||\xrightarrow{p}0$. Thus, $||\hat{\theta}_n-\theta_0||\xrightarrow{p}0$. ■

After showing consistency we move our attention to the asymptotic normality of the estimator. In order to derive the limiting distribution define

$$V_{1\theta}=E[\psi_\theta(y, x)\psi_\theta(y, x)'], \tag{22}$$

and

$$V_{2\theta}=\frac{\partial E[\psi_\theta(y, x)]}{\partial\theta'}. \tag{23}$$

Here,

$$V_{1\theta}=\begin{bmatrix} \frac{1}{\sigma}\tau(1-\tau)E[xx'] & \frac{E[((1-2\tau)\text{sign}(y-x'\beta)-(1-2\tau)^2)x']}{2\sigma\tau(1-\tau)}-E\left[\frac{1}{\sigma^2}\rho_\tau(y-x'\beta)x'\right] & \frac{1}{2\sigma^3}E[\rho_\tau(y-x'\beta)(\text{sign}(y-x'\beta)-(1-2\tau))x'] \\ . & \frac{(1-2\tau)^2}{\tau^2(1-\tau)^2}+E\left[\frac{1}{\sigma^2}(y-x'\beta)^2\right]-2\frac{(1-2\tau)}{\tau(1-\tau)}E\left[\frac{1}{\sigma}(y-x'\beta)\right] & \frac{1}{\sigma^2}E\left[\rho_\tau(y-x'\beta)\left(\frac{(1-2\tau)}{\tau(1-\tau)}-\frac{1}{\sigma}(y-x'\beta)\right)\right] \\ . & . & \frac{1}{\sigma^4}E[\rho_\tau^2(y-x'\beta)]+\frac{1}{\sigma^2}-\frac{1}{\sigma^3}E[\rho_\tau(y-x'\beta)] \end{bmatrix}$$

and

$$V_{2\sigma}=\begin{bmatrix} -\dfrac{E[g(x'\beta|x)xx']}{\sigma} & \dfrac{1}{\sigma}E[x] & 0 \\ \cdot & \dfrac{-1+2\tau-2\tau^2}{\tau^2(1-\tau)^2} & \dfrac{1}{\sigma^2}E[(y-x'\beta)] \\ \cdot & \cdot & -\dfrac{1}{\sigma^2} \end{bmatrix}.$$

Note that when $y|x\sim ALPD(x'\beta, \tau, \sigma)$, then $V_{1\theta}=-V_{2\theta}$.

Assumption A6 Assume that $V_{1\theta_0}$ and $V_{2\theta_0}$ exist and are finite, $V_{1\theta_0}$ is positive definite and $V_{2\theta_0}$ is invertible.

Chernozhukov, Fernández-Val, and Melly (2009) calculated equations (22) and (23) in the quantile process as an approximate Z-estimator.

Now we state the asymptotic normality result.

Theorem 2 *Under Assumptions 1–6,*

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\Rightarrow N(0,\ V_{2\theta_0}^{-1}V_{1\theta_0}V_{2\theta_0}^{-1}).$$

Proof: First, combining Theorem 1 and second part of Lemma A1, we have

$$\mathbb{G}_n\psi_{\hat{\theta}_n}(y, x)=\mathbb{G}_n\psi_{\theta_0}(y, x)+o_p(1),$$

where $f\mapsto\mathbb{G}_n[f(w)]=\dfrac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(w_i)-Ef(w_i))$. Rewriting we have

$$\sqrt{n}\mathbb{E}_n\psi_{\hat{\theta}_n}(y, x)=\sqrt{n}E\psi_{\hat{\theta}_n}(y, x)+\mathbb{G}_n\psi_{\theta_0}(y, x)+o_p(1). \tag{24}$$

By Assumption A5

$$||\mathbb{E}_n\psi_{\hat{\theta}_n}(y, x)||=o_p(n^{-1/2}) \text{ and } E[\psi_{\theta_0}(y, x)]=0.$$

Now consider the first element of the right hand side of (24). By a Taylor expansion about $\hat{\theta}_n=\theta_0$ we obtain

$$E[\psi_{\hat{\theta}_n}(y, x)]=E[\psi_{\theta_0}(y, x)]+\frac{\partial E[\psi_\theta(y, x)]}{\partial\theta'}\Big|_{\theta=\theta_0}(\hat{\theta}_n-\theta_0)+o_p(1), \tag{25}$$

where

$$\frac{\partial E[\psi_\theta(y, x)]}{\partial\theta'}\Big|_{\theta=\theta_0}=\frac{\partial}{\partial\theta'}E\begin{pmatrix} \dfrac{1}{\sigma}(\tau-1(y<x'\beta))x \\ \left(\dfrac{1-2\tau}{\tau(1-\tau)}-\dfrac{(y-x'\beta)}{\sigma}\right) \\ -\dfrac{1}{\sigma}+\dfrac{1}{\sigma^2}\rho_\tau(y-x'\beta) \end{pmatrix}\Big|_{\theta=\theta_0}.$$

Since by condition A6, $\dfrac{\partial E[\psi_\theta(y, x)]}{\partial\theta'}\Big|_{\theta=\theta_0}=V_{2\theta_0}$, equation (25) can be rewritten as

$$E[\psi_\theta(y, x)]|_{\theta=\hat{\theta}_n}=V_{2\theta_0}(\hat{\theta}_n-\theta_0)+o_p(1). \tag{26}$$

Using Assumption A5 (ii), from (24) we have

$$o_p(1)=\sqrt{n}E\psi_{\hat{\theta}_n}(y, x)+\mathbb{G}_n\psi_{\theta_0}(y, x)+o_p(1),$$

and using the above approximation given in (26)

$$o_p(1)=V_{2\theta_0}\sqrt{n}(\hat{\theta}_n-\theta_0)+\mathbb{G}_n\psi_{\theta_0}(y,x)+o_p(1).$$

By invertibility of $V_{2\theta_0}$ in A6,

$$\sqrt{n}(\hat{\theta}_n-\theta_0)=-V_{2\theta_0}^{-1}\mathbb{G}_n\psi_{\theta_0}(y,x)+o_p(1). \tag{27}$$

Finally, from Lemma A1 $\theta\mapsto\mathbb{G}_n\psi_\theta(y,x)$ is stochastic equicontinuous. So, stochastic equicontinuity and ordinary CLT imply that $\mathbb{G}_n\psi_\theta(y,x)\Rightarrow z(\cdot)$ converges to a Gaussian process with variance-covariance function defined by $V_{1\theta_0}=E[\psi_\theta(y,x)\psi_\theta(y,x)']|_{\theta=\theta_0}$. Therefore, from (27)

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\Rightarrow V_{2\theta_0}^{-1}z(\cdot),$$

so that

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\Rightarrow N(0,\ V_{2\theta_0}^{-1}V_{1\theta_0}V_{2\theta_0}^{-1}).$$ ■

4 Monte Carlo Simulations

In this section we provide a glimpse into the finite sample behavior of the proposed ZQR estimator. Two simple versions of our basic model are considered in the simulation experiments. In the first, results of which are reported in Table 1, the scalar covariate, x_i, exerts a pure location shift effect. In the second, reported in Table 2, x_i has both a location and scale shift effects. In the former case the response, y_i, is generated by the model,

$$y_i=\alpha+\beta x_i+u_i,$$

while in the latter case,

$$y_i=\alpha+\beta x_i+(1+\gamma x_i)u_i,$$

Table 1: Location-Shift Model: Bias and RMSE.

	ZQR	LAD	OLS
$N(0,1)$			
Bias	0.0007	−0.0004	0.0008
RMSE	0.0904	0.0899	0.0710
$\hat{\tau}$	0.501	–	–
t_3			
Bias	0.0012	−0.0008	0.0014
RMSE	0.1133	0.0967	0.1217
$\hat{\tau}$	0.498	–	–
χ_3^2			
Bias	−0.0021	0.0024	0.0020
RMSE	0.1419	0.1892	0.1801
$\hat{\tau}$	0.081	–	–
ALPD (τ=0.5)			
Bias	0.0001	0.0001	0.0001
RMSE	0.0638	0.0549	0.0710
$\hat{\tau}$	0.499	–	–
ALPD (τ=0.25)			
Bias	−0.0008	−0.0001	0.0003
RMSE	0.0718	0.0860	0.0917
$\hat{\tau}$	0.248	–	–

Table 2: Location-Scale-Shift Model: Bias and RMSE.

	ZQR	LAD	OLS
$N(0, 1)$			
Bias	0.0015	0.0036	0.0037
RMSE	0.2209	0.1461	0.1365
$\hat{\tau}$	0.499	–	–
t_3			
Bias	−0.0005	0.0002	−0.0052
RMSE	0.2457	0.1460	0.2565
$\hat{\tau}$	0.501	–	–
χ^2_3			
Bias	−0.0004	0.0076	0.0089
RMSE	0.5087	0.2833	0.3788
$\hat{\tau}$	0.086	–	–
ALPD (τ=0.5)			
Bias	−0.0010	−0.0001	−0.0013
RMSE	0.1455	0.0845	0.1459
$\hat{\tau}$	0.501	–	–
ALPD (τ=0.25)			
Bias	0.0051	0.0004	0.4076
RMSE	0.1331	0.1429	0.4505
$\hat{\tau}$	0.248	–	–

where u_i are i.i.d. innovations generated according to a standard normal distribution, t_3 distribution, χ^2_3 centered at the mean, Laplace distribution (i.e., τ=0.5), and ALPD with τ=0.25.[3] In the location shift model x_i follows a standard normal distribution; in the location-scale shift model, it follows a χ^2_3. We set $\alpha=\beta=1$ and γ=0.5. Our interest is on the effect of the covariates in terms of bias and root mean squared error (RMSE). We carry out all the experiments with sample size n=200 and 5000 replications. Three estimators are considered: our proposed ZQR estimator, QR at the median, i.e., the least-absolute deviation (LAD) estimator, and ordinary least squares (OLS). We pay special attention to the estimated quantile $\hat{\tau}$ in the ZQR.

Note that there are fundamental differences between the location and location-scale models. First, for the location model, ZQR, LAD and OLS should be estimating β=1, and both bias and RMSE are constructed by comparing $\hat{\beta}_j$, $j=ZQR, LAD, OLS$ with 1. Second, for the location-scale model our proposed ZQR, LAD and OLS estimators are not directly comparable in all cases. The reason is that the parameter value to which each estimator should be compared for constructing bias and RMSE may differ. In particular, in the location-scale model, the value of the parameter to be estimated depends on the selected quantile, that is, $\beta(\tau)=\beta+\gamma Q_u(\tau)$, for LAD and ZQR but not for OLS (where β is the parameter to be estimated). For LAD this is $\beta(0.5)=\beta+\gamma Q_u(0.5)$. For ZQR a different τ should be used in $\beta(\tau)=\beta+\gamma Q_u(\tau)$ depending on the distribution because in this case τ should be set at the most probable quantile. In particular, τ=0.5 for $N(0, 1)$, t_3 and $ALPD(\tau=0.5)$, but τ=0.25 for $ALPD(0.25)$ and τ=0.085 for χ^2_3.

Table 1 reports the results for the location shift model. In all cases we compute the bias and RMSE with respect to β=1. Bias is close to zero in all cases. In the Gaussian setting, as expected, we observe efficiency loss in ZQR and QR estimates compared to that of OLS. Under symmetric distributions, normal, t_3, and Laplace, the estimated quantile of interest $\hat{\tau}$ in the ZQR is remarkably close to 0.5. In the χ^2_3 case, the ZQR estimator performs better than the QR and OLS procedures. Note that the estimated quantile for the χ^2_3 is 0.081, consistent with the fact that the underline distribution is positively skewed. Finally, for the ALPD(0.25) case,

3 Although not reported, similar results were obtained for ALPD with τ=0.75.

ZQR produces the estimated quantile ($\hat{\tau}$=0.248) very close to 0.25, and also has a smaller RMSE. Overall, Table 1 shows that the ZQR estimator retains the robustness properties of the QR estimator, although we do not specify a particular quantile of interest.

In the location-scale version of the model we adopt the same data generating mechanism. For this case the effect of the covariate x_i on quantile of interest response in QR is given by $\beta(\tau)=\beta+\gamma Q_u(\tau)$. In ZQR we compute bias and RMSE by averaging estimated τ from 5000 replications. The results are reported in Table 2. The results for the normal, t_3 and Laplace distributions are similar to those in the location model, showing that all point estimates are approximately unbiased. As expected, OLS outperforms ZQR and QR in the normal case, but the opposite occurs in the t_3 and Laplace distributions. In the χ^2_3 case, the estimated quantile is $\hat{\tau}$=0.086, very close to the value of 0.085. For the ALPD(0.25) distribution, the best performance is also obtained for the ZQR estimator.

5 Empirical Illustration: The Effect of Job Training on Wages

The effect of policy variables on distributional outcomes are of fundamental interest in empirical economics. Of particular interest is the estimation of the quantile treatment effects (QTE), that is, the effect of some policy variable of interest on the different quantiles of a response variable. Our proposed estimator complements the QTE analysis by providing a parsimonious estimator at the most probable quantile.

We apply the estimator to the study of the effect of public-sponsored training programs. As argued in LaLonde (1995), public programs of training and employment are designed to improve participant's productive skills, which in turn would increase their earning potential and decrease dependency on social welfare benefits. We use data from the Job Training Partnership Act (JTPA) program, that has been extensively studied in the literature. For example, see Bloom et al. (1997) for a description, and Abadie, Angrist, and Imbens (2002) for QTE analysis. The JTPA was a large publicly-funded training program that began funding in October 1983 and continued until late 1990s. We focus on the Title II subprogram, which was offered only to individuals with "barriers to employment" (long-term use of welfare, being a high-school drop-out, 15 or more recent weeks of unemployment, limited English proficiency, phsysical or mental disability, reading proficiency below 7th grade level or an arrest record). Individuals in the randomly assigned JTPA treatment group were offered training, while those in the control group were excluded for a period of 18 months. Our interest lies in measuring the effect of a training offer and actual training on of participants' future earnings.

We use the database in Abadie, Angrist, and Imbens (2002) that contains information about adult male and female JTPA participants and non-participants. Let z denote the indicator variable for those receiving a JTPA offer. Of those offered, 60% completed the training; of those in the control group completion rate was <2%. For illustrating the use of ZQR, we first study the effect of receiving a JTPA offer on log wages, and later we pursue instrumental variables estimation in the ZQR context. Following Abadie, Angrist, and Imbens (2002) we use a linear regression specification model, where the JTPA offer enters in the equation as a dummy variable.[4] We consider the following regression model:

$$y=z\gamma+x\beta+u,$$

where the dependent variable y is the logarithm of 30 month accumulated earnings (we exclude individuals without earnings), z is a dummy variable for the JTPA offer, x is a set of exogenous covariates containing individual characteristics, and u is an unobservable component. The parameter of interest is γ that provides the effect of training offer on wages.

First, we compute the QR process for all $\tau\in$ (0.05, 0.95) and the results are presented in Figure 2. The JTPA effect estimates for QR and OLS appear in Table 3. Interestingly, with exception of low quantiles, the effect

4 Linear regression models are common in the QTE literature to accomodate several control variables capturing individual characteristics. See for instance Chernozhukov and Hansen (2006, 2008) and Firpo (2007).

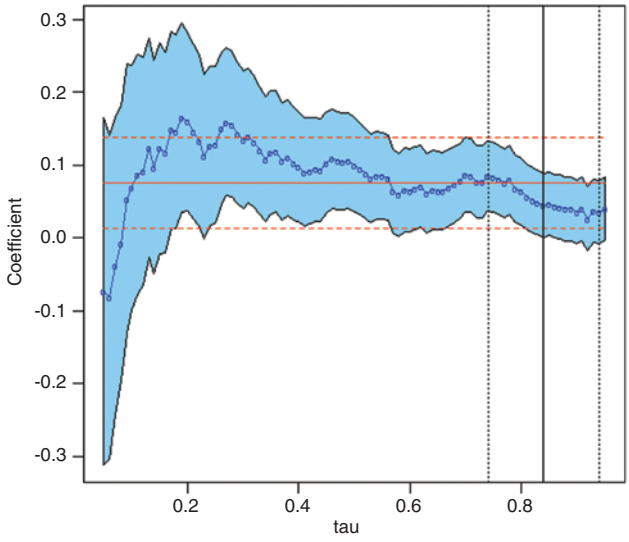

Figure 2: JTPA offer: Quantile regression process and OLS.
Notes: Quantile regression process (shaded area), OLS (horizonal lines) and estimated most informative quantile (vertical lines) with 95% confidence intervals.

of JTPA is decreasing in τ, which implies that the individuals in the high quantiles of the conditional wage distribution benefited less from the JTPA training. Second, by solving equation (19) we obtain $\hat{\tau}$=0.84 and this means that the distribution of unobservables is negatively skewed and quantiles to the left of the median are a better summary of the data. Uniqueness of the selected quantile is checked through Figure 3, where $\hat{\tau}$ is indicted by a vertical solid line, together with the 95% confidence interval given by the vertical dotted lines.

From Table 3 we observe that the training effect estimate from mean and median regressions are, respectively, 0.075 (0.032) and 0.100 (0.033) which are similar, however they both are larger than the ZQR estimate of 0.045 (0.022), the numbers in parenthesis are the corresponding standard errors. Figure 2 shows that QR

Table 3: JTPA offer.

	ZQR [$\hat{\tau}$=0.84]	OLS	LAD
Intercept	9.894 (0.059)	8.814 (0.088)	9.188 (0.086)
JTPA offer	0.045 (0.022)	0.075 (0.032)	0.100 (0.033)
Female	0.301 (0.023)	0.259 (0.030)	0.260 (0.031)
HSORGED	0.201 (0.025)	0.267 (0.034)	0.297 (0.037)
Black	−0.102 (0.026)	−0.121 (0.036)	−0.175 (0.039)
Hispanic	−0.032 (0.034)	−0.034 (0.050)	−0.025 (0.051)
Married	0.129 (0.025)	0.242 (0.036)	0.265 (0.034)
WKLESS13	−0.255 (0.023)	−0.598 (0.032)	−0.556 (0.036)
Age2225	0.229 (0.057)	0.175 (0.084)	0.125 (0.080)
AGE2629	0.285 (0.058)	0.192 (0.085)	0.131 (0.081)
AGE3035	0.298 (0.057)	0.191 (0.084)	0.176 (0.080)
AGE3644	0.320 (0.058)	0.130 (0.085)	0.173 (0.081)
AGE4554	0.267 (0.064)	0.110 (0.094)	0.080 (0.092)
$\hat{\tau}$	0.840 (0.051)		0.500
$\hat{\sigma}$	0.249 (0.060)		0.538 (0.006)

9872 observations. The numbers in parenthesis are the corresponding standard errors. JTPA offer, dummy variable for individuals that received a JTPA offer; FEMALE, Female dummy variable; HSORGED, dummy variable for individuals with completed high school or GSE; BLACK, race dummy variable; HISPANIC, dummy variable for hispanic; MARRIED, dummy variable for married individuals; WKLESS13, dummy variable for individuals working <13 weeks in the past year; AGE2225, AGE2629, AGE3035, AGE3644 and AGE4554 age range indicator variables.

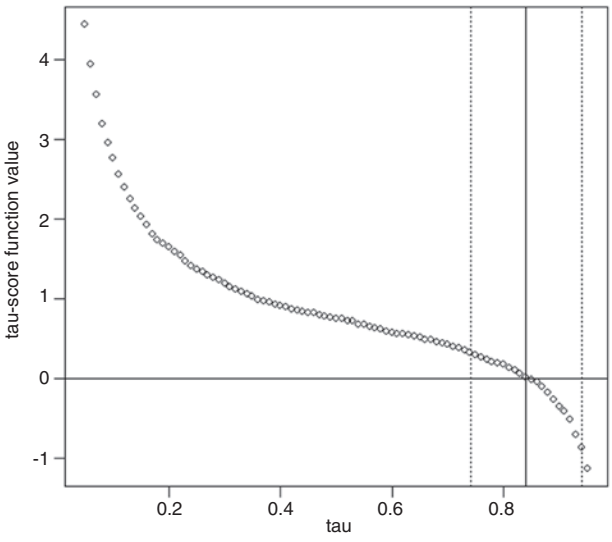

Figure 3: JTPA offer: τ-score function.

Notes: The τ-score function is $\frac{1-2\tau}{\tau(1-\tau)}-\frac{\sum_{i=1}^{n}(y_i-x_i'\hat{\beta}(\tau))}{n\hat{\sigma}}$.

estimates in the upper tail of the distribution have smaller standard errors, which suggests that by choosing the most likely quantile the ZQR procedure implicitly solves for the smallest standard error QR estimator. The results show that for the ZQR quantile, $\hat{\tau}=0.84$ (0.051), the effect of training is less than the mean and median effects. From a policy maker perspective, if one is asked to report the effect of training on wage, it could be done through the mean effect (0.075), the median effect (0.100) or even the entire conditional quantile function as in Figure 2; our analysis suggests reporting the most likely effect (0.045) coming from the representative quantile $\hat{\tau}=0.84$. Using the above model, the fit of the data reveals that the upper quantiles are informative, and the ZQR estimator is appropriate to describe the effect of JTPA on earnings.

As argued in the Introduction, the ZQR framework allows for a different interpretation of the QR analysis. Suppose that we are interested in a targeted treatment effect of $\bar{\gamma}=0.1$, and we would like to get the representative quantile of the unobservables distribution that will most likely have this effect. This corresponds to estimating the ZQR parameters for $\gamma-z\bar{\gamma}=x\beta+u$. In this case, we obtain an estimated quantile of $\widehat{\tau(\bar{\gamma})}=0.85$. Note that $\hat{\sigma}=0.249$ for ZQR while $\hat{\sigma}=0.538$ for LAD, which implies that ZQR has a much better goodness-of-fit than LAD.

To value the option of treatment is an interesting exercise in itself, but policy makers may be more interested in the effect of actual training rather than the possibility of training. In this case the model of interest is

$$y=d\alpha+x\beta+u,$$

where d is a dummy variable indicating if the individual actually completed the JTPA training. We have strong reasons to believe that $cov(d, u)\neq 0$ and therefore, OLS and QR estimates will be biased. In this case, while the JTPA offer is random, those individuals who decide to undertake training do not constitute a random sample of the population. Rather, they are likely to be more motivated individuals or those who view the training to be valuable. However, the exact nature of this bias is unknown in terms of quantiles. Figure 4 reports the entire quantile process and OLS for the above equation. Interestingly the effect of training on wages is monotonically decreasing in τ. The selection of the most likely quantile determines that as in the previous case $\hat{\tau}=0.84$.

In order to solve the potential endogeneity problem, following Abadie, Angrist, and Imbens (2002), z can be used as a valid instrument for d. The reason is that it is exogenous as it was a randomized experiment, and it is correlated with d (as mentioned earlier 60% of individuals undertook training when they were offered).

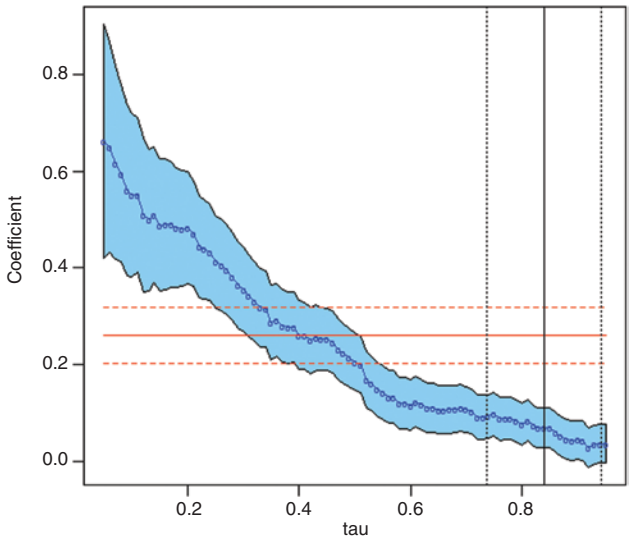

Figure 4: JTPA: Quantile regression process and OLS.
Notes: Quantile regression process (shaded area), OLS (horizonal lines) and estimated most informative quantile (solid vertical line) with 95% confidence intervals (dotted lines).

The instrumental variable (IV) strategy is based on Chernozhukov and Hansen (2006, 2008) by considering the model

$$y-d\alpha=x\beta+z\gamma+u.$$

The IV method in QR proceeds as follows. Note that z does not belong to the model, as conditional on d, undertaking training, the offer has no effect on wages. Then, we construct a grid in $\alpha\in\mathcal{A}$, which is indexed by j for each $\tau\in(0,1)$ and we estimate the quantile regression model for fixed τ

$$y-d\alpha_j(\tau)=x\beta+z\gamma+u.$$

This gives $\{\hat{\beta}_j(\alpha_j(\tau),\tau),\hat{\gamma}_j(\alpha_j(\tau),\tau)\}$, the set of conditional quantile regression estimates for the new model. Next, we choose α by minimizing a given norm of γ (we use the Euclidean norm),

$$\hat{\hat{\alpha}}(\tau)=argmin_{\alpha\in\mathcal{A}}\|\hat{\gamma}(\alpha(\tau),\tau)\|.$$

Figure 5 depicts the values of γ^2 for the grids of α and τ. As a result we obtain the map $\tau\mapsto\{\hat{\hat{\alpha}}(\tau),\hat{\beta}(\hat{\hat{\alpha}}(\tau),\tau)\equiv\hat{\hat{\beta}}(\tau),\hat{\gamma}(\hat{\hat{\alpha}}(\tau),\tau)\equiv\hat{\hat{\gamma}}(\tau)\}$.

Finally, we select the most probably quantile as in the previous case, by using the first order condition corresponding the selection of τ:

$$\hat{\hat{\tau}}=argmin_{\tau\in(0,1)}\left|\frac{1-2\tau}{\tau(1-\tau)}-\frac{\sum_{i=1}^{n}\hat{\hat{u}}_i(\tau)}{\sum_{i=1}^{n}\rho_\tau(\hat{\hat{u}}_i(\tau))}\right|,$$

where $\hat{\hat{u}}_i(\tau)=y_i-d_i\hat{\hat{\alpha}}(\tau)-x_i'\hat{\hat{\beta}}(\tau)-z_i\hat{\hat{\gamma}}(\tau)$. Figure 6 reports the IV estimates together with the most likely quantile. Interestingly, the qualitative results are very much alike those for the JTPA training offer. The IV least-squares estimator for the effect of JTPA training gives a value of 0.116 (0.045) while IV median regression gives a much higher value of 0.142 (0.047). The most likely quantile is still the same 0.84 (0.053), with an associated training effect of 0.072 (0.033). Thus, the ZQR effect continues to be smaller than the mean and median

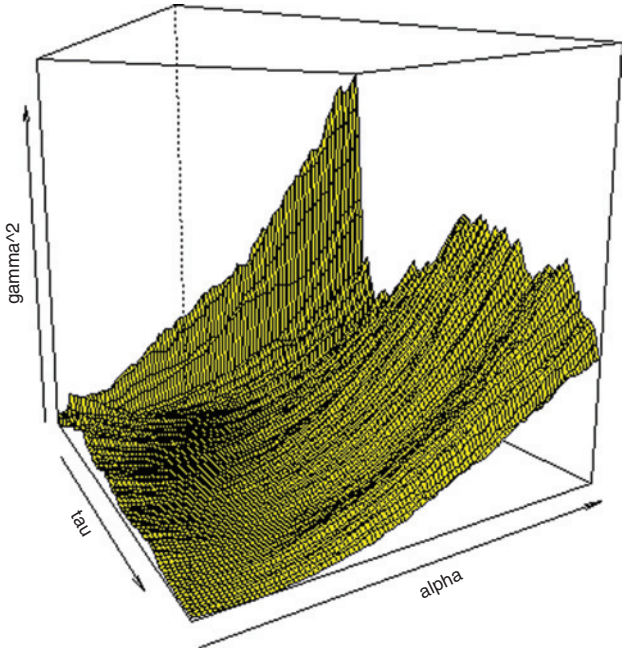

Figure 5: JTPA: Minimization of $||\gamma^2(\tau, \alpha)||$.

estimates. We notice that the most likely impact to be in the upper quantiles when analyzing the effects of JTPA training on log wages.

6 Conclusions

In this paper we show that the maximum likelihood problem for the asymmetric Laplace distribution can be found as the solution of a maximum entropy problem where we impose moment constraints given by the joint

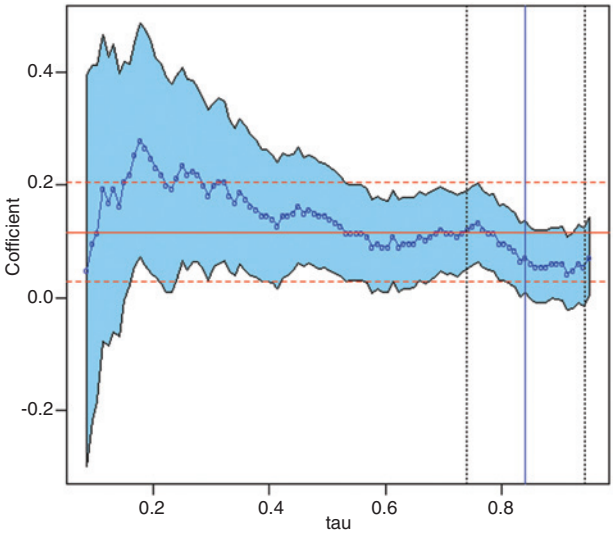

Figure 6: JTPA: IV Quantile regression process and IV OLS.
Notes: Quantile regression process (shaded area), OLS (horizonal lines) and estimated most informative quantile (solid vertical line) with 95% confidence intervals (dotted lines).

consideration of the mean and median. We also propose an approximate Z-estimator method, which provides a parsimonious estimator that complements the quantile process. This provides an alternative interpretation of quantile regression and puts it within the likelihood framework and maximum entropy paradigm. Potential estimates from this method have practical implications. As an illustration, we apply the proposed estimator to a well-known dataset from a job training program where quantile regression has been extensively used.

As noted by an anonymous referee our proposed ZQR estimator can be revised as a flexible MLE with skewed distributions (e.g., skewed normal or skewed Student's t). The ZQR estimator also inherits some of the robustness properties of QR and should be thus considered a complement to the quantile analysis. However, it has the non-differentiability problem of QR which makes the asymptotic analysis more challenging than the standard normal or Student's MLE framework.

Acknowledgments: We are very grateful to the Editor, two anonymous referees, Arnold Zellner, Jushan Bai, Rong Chen, Daniel Gervini, Yongmiao Hong, Carlos Lamarche, Ehsan Soofi, Liang Wang, Zhijie Xiao, and the participants in seminars at University of Wisconsin-Milwaukee, City University London, Info-Metrics Institute Conference, September 2010, World Congress of the Econometric Society, Shanghai, August 2010, Latin American Meeting of the Econometric Society, Argentina, October 2009, Summer Workshop in Econometrics, Tsinghua University, Beijing, China, May 2009, for helpful comments and discussions. However, we retain the responsibility for any remaining errors.

Appendix

A. Interpretation of the Z-estimator

In order to interpret θ_0, we take the expectation of the estimating equations with respect to the unknown true density. To simplify the exposition we consider a simple model without covariates: $y_i=\alpha+u_i$. Our estimating equation vector is defined as:

$$E(\Psi_\theta(y))=E\begin{pmatrix}\frac{1}{\sigma}(\tau-1(y<\alpha))\\ \left(\frac{1-2\tau}{\tau(1-\tau)}-\frac{(y-\alpha)}{\sigma}\right)\\ -\frac{1}{\sigma}+\frac{1}{\sigma^2}\rho_\tau(y-\alpha)\end{pmatrix}=0,$$

and the estimator is such that

$$\frac{1}{n}\sum_{i=1}^{n}\psi_\theta(y_i)=0$$

Let $F(y)$ be the cdf of the random variable y. Now we need to find $E[\Psi_\theta(y)]$.

For the first component we have

$$\begin{aligned}\frac{1}{\sigma}E[\tau-I(y<\alpha)]&=\frac{1}{\sigma}\left(\int_{\mathbb{R}}(\tau-1(y<\alpha))dF(y)\right)\\&=\frac{1}{\sigma}\left(\tau-\int_{-\infty}^{\alpha}dF(y)\right)\\&=\frac{1}{\sigma}(\tau-F(\alpha)).\end{aligned}$$

Thus if we set this equal to zero, we have

$$\alpha=F^{-1}(\tau),$$

which is the usual quantile. Thus, the interpretation of the parameter α is analogous to QR if covariates are included.

For the third term in the vector, $-\frac{1}{\sigma}+\frac{1}{\sigma^2}\rho_\tau(y-\alpha)$, we have

$$E\left[-\frac{1}{\sigma}+\frac{1}{\sigma^2}\rho_\tau(y-\alpha)\right]=0,$$

that is,

$$\sigma=E[\rho_\tau(y-\alpha)].$$

Thus, as in the least squares case, the scale parameter σ can be interpreted as the expected value of the loss function.

Finally, we can interpret τ using the second equation,

$$E\left[\frac{1-2\tau}{\tau(1-\tau)}-\frac{(y-\alpha)}{\sigma}\right]=0,$$

which implies that

$$\frac{1-2\tau}{\tau(1-\tau)}=\frac{E[y]-F^{-1}(\tau)}{\sigma}.$$

Note that $s(\tau)\equiv\frac{1-2\tau}{\tau(1-\tau)}$ is a measure of the skewness of the distribution. Thus, τ should be chosen to set $s(\tau)$ equal to a measure of asymmetry of the underline distribution $F(\cdot)$ given by the difference of τ-quantile with the mean (and standardized by σ). In the special case of a symmetric distribution, the mean coincides with the median and mode, such that $E[y]=F^{-1}(1/2)$ and $\tau=1/2$, which is the most probable quantile and a solution to our Z-estimator.

B. Lemma A1

In this appendix we state an auxiliary result that states Donskerness and stochastic equicontinuity. Let $\mathcal{F}\equiv\{\psi_\theta(y, x), \theta\in\Theta\}$, and define the following empirical process notation for $w=(y, x)$:

$$f\mapsto\mathbb{E}_n[f(w)]=\frac{1}{n}\sum_{i=1}^{n}f(w_i)\quad f\mapsto\mathbb{G}_n[f(w)]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(w_i)-Ef(w_i)).$$

We follow the literature using empirical process exploiting the monotonicity and boundedness of the indicator function, the boundedness of the moments of x and y, and that the problem is a parametric one.

Lemma A1 Under Assumptions A1–A4 $\mathcal{F}$ is Donsker. Furthermore,

$$\theta\mapsto\mathbb{G}_n\psi_\theta(y, x)$$

is stochastically equicontinuous, that is

$$\sup_{||\theta-\theta_0||\le\delta_n}||\mathbb{G}_n\psi_\theta(y, x)-\mathbb{G}_n\psi_{\theta_0}(y, x)||=o_p(1),$$

for any $\delta_n\downarrow 0$.

Proof: The proof of this result follows similar steps to those in Chernozhukov and Hansen (2006). To prove the lemma we check the conditions for independent but not identically distributed process stated in Theorem 2.11.1 of van der Vaart and Wellner (1996). It is important to note that a class $\mathcal{F}$ of a vector-valued functions $f:x\mapsto\mathbb{R}^k$ is Donsker if each of the classes of coordinates $f_j:x\mapsto\mathbb{R}^k$ with $f=(f_1, \ldots, f_k)$ ranging over $\mathcal{F}(j=1, 2,\ldots,k)$ is Donsker (van der Vaart 1998, 270).

First, one can check the random-entropy condition by checking that $\mathcal{F}$ satisfies a uniform entropy condition and the envelope is square integrable. The first element of the vector is $\psi_{1\theta}(y,x)=(\tau-1(y<x'\beta))\frac{x}{\sigma}$. Note that the functional class $\mathfrak{N}=\{\tau-1\{y<x'\beta\}, \tau\in\mathcal{T}, \beta\in\mathcal{B}\}$ is a VC subgraph class, with envelope 2. Its product with x also forms a class with a square integrable envelope $2\max_j|x_j|$. Finally, the class $\mathcal{F}_1$ is defined as the product of the latter with $1/\sigma$, which is bounded by assumption A3. Thus, by assumption A4 $\mathcal{F}_1$ is satisfies a uniform entropy condition and the envelope function is square integrable. Therefore, the random entropy condition (2.11.2) in van der Vaart and Wellner (1996) is satisfied.

The second element of the vector is $\psi_{2\theta}(y,x)=\left(\frac{1-2\tau}{\tau(1-\tau)}-\frac{(y-x'\beta)}{\sigma}\right)$. Define $\mathfrak{H}=\{(y-x'\beta), \beta\in\mathcal{B}\}$. Note that

$$|(y-x'\beta_1)-(y-x'\beta_2)|=|x'(\beta_2-\beta_1)|\le||x||||\beta_2-\beta_1||,$$

where the inequality follows from Cauchy-Schwartz inequality. Thus by Assumptions A3–A4 the class $\mathfrak{H}$ has envelope function square integrable. In addition, note that, $\mathfrak{H}$ belongs to a VC class satisfying a uniform entropy condition, since this class is a subset of the vector space of functions spanned by $(y, x_1, \ldots, x_p)$, where p is the fixed dimension of x (see e.g., example 19.7 in van der Vaart (1998)). Thus, the class defined by $1/\sigma\,\mathfrak{H}$ has envelope of $\mathfrak{H}$ ($|y|+const*|x|$) which is square integrable by assumptions A3–A4. Therefore, $\mathcal{F}_2$ satisfies the random entropy condition.

The third element of the vector is $\psi_{3\theta}(y,x)=\left(-\frac{1}{\sigma}+\frac{1}{\sigma^2}\rho_\tau(y-x'\beta)\right)$. Consider the following random process defined by $\mathfrak{I}=\{\rho_\tau(y-x'\beta), \tau\in\mathcal{T}, \beta\in\mathcal{B}\}$ which satisfies a uniform entropy condition and square integrable envelope function. The latter is given by Assumptions A3–A4 and the quantile regression check function properties as $\rho_\tau(x+y)-\rho_\tau(y)\le 2|x|$ and $\rho_{\tau_1}(y-x't)-\rho_{\tau_2}(y-x't)=(\tau_2-\tau_1)(y-x't)$. The former follows from the fact that this is a parametric class collection of measurable functions indexed in a bounded subset. Hence, $\mathcal{F}_3$ the random entropy condition and the envelope function is square integrable.

Now we turn our attention to the second condition of Theorem 2.11.1 in van der Vaart and Wellner (1996). The process $\theta\mapsto\mathbb{G}_n\psi_\theta(y,x)$ is stochastically equicontinuous over Θ with respect to a $L_2(P)$ pseudo-metric. First, as in Angrist, Chernozhukov, and Fernández-Val (2006) and Chernozhukov and Hansen (2006), we define the distance d as the following $L_2(P)$ pseudo-metric

$$d(\theta',\theta'')=\sqrt{E([\psi_{\theta'}-\psi_{\theta''}]^2)}.$$

Thus, as $||\theta-\theta_0||\to 0$ we need to show that

$$d(\theta,\theta_0)\to 0, \tag{28}$$

and the final follows from Theorem 2.11.1 of van der Vaart and Wellner (1996).

To show (28), first note that for each $i=1,\ldots,n$,

$$\begin{aligned}
d_{1i}(\theta',\theta)&=\sqrt{E([\psi_{1\theta'}-\psi_{1\theta}]^2)}\\
&=\sqrt{E\left(\left[(\tau'-1(y_i-x_i\beta'))\frac{x_i}{\sigma'}-(\tau-1(y_i-x_i\beta))\frac{x_i}{\sigma}\right]^2\right)}\\
&\le\left[\left(E\left|\frac{1}{\sigma'}(\tau'-1(y_i-x_i\beta'))-\frac{1}{\sigma}1(\tau-(y_i-x_i\beta))\right|^{\frac{2(2+\epsilon)}{\epsilon}}\right)^{\frac{\epsilon}{(2+\epsilon)}}\cdot\left(E(|x_i|^2)^{\frac{2+\epsilon}{2}}\right)^{\frac{2}{(2+\epsilon)}}\right]^{\frac{1}{2}}\\
&=\left(E\left|\left(\frac{\tau'}{\sigma'}-\frac{\tau}{\sigma}\right)+\left(\frac{1}{\sigma}1(y_i\le x_i\beta)-\frac{1}{\sigma'}1(y_i\le x_i\beta')\right)\right|^{\frac{2(2+\epsilon)}{\epsilon}}\right)^{\frac{\varepsilon}{2(2+\epsilon)}}\cdot\left(E(|x_i|^2)^{\frac{2+\epsilon}{2}}\right)^{\frac{1}{(2+\epsilon)}}
\end{aligned}$$

$$\leq \left[\left(\left(\left| \frac{\tau'}{\sigma'} - \frac{\tau}{\sigma} \right| \right)^{\frac{2(2+\epsilon)}{\epsilon}} \right)^{\frac{\epsilon}{2(2+\epsilon)}} + \left(E\left(\left| \frac{1}{\sigma} 1(y_i \leq x_i\beta) - \frac{1}{\sigma'} 1(y_i \leq x_i\beta') \right| \right)^{\frac{2(2+\epsilon)}{\epsilon}} \right)^{\frac{\epsilon}{2(2+\epsilon)}} \right] \cdot \left(E(|x_i|^2)^{\frac{2+\epsilon}{2}} \right)^{\frac{1}{(2+\epsilon)}}$$

$$\leq \left[\left| \frac{\tau'}{\sigma'} - \frac{\tau}{\sigma} \right| + \left(E\left| \overline{g}_i \cdot x_i \left(\frac{\beta'}{\sigma'} - \frac{\beta}{\sigma} \right) \right| \right)^{\frac{\epsilon}{2(2+\epsilon)}} \right] \cdot (E||x_i||^{2+\epsilon})^{\frac{1}{(2+\epsilon)}}$$

$$\leq \left[\left| \frac{\tau'}{\sigma'} - \frac{\tau}{\sigma} \right| + \left(E\,\overline{g}_i ||x_i|| \left| \left| \frac{\beta'}{\sigma'} - \frac{\beta}{\sigma} \right| \right| \right)^{\frac{\epsilon}{2(2+\epsilon)}} \right] \cdot (E||x_i||^{2+\epsilon})^{\frac{1}{(2+\epsilon)}},$$

where the first inequality is Holder's inequality, the second is Minkowski's inequality, the third is a Taylor expansion as in Angrist, Chernozhukov and Fernández-Val (2006), p.560) where $\overline{g}_i$ is the upper bound of $g_i(y_i|x)$ (using A2), and the last is Cauchy-Schwarz inequality. Therefore, by assumption A2–A4 $\sup_{||\theta'-\theta||<\delta_n} \sum_{i=1}^{n} d_{1i} \to 0$ when $\delta_n \to 0$.

Now rewrite $\psi_{2\theta}(y, x) = \left(\sigma \frac{1-2\tau}{\tau(1-\tau)} - (y - x'\beta) \right)$ and note that

$$d_{2i}(\theta', \theta) = \sqrt{E([\psi_{2\theta'} - \psi_{2\theta}]^2)}$$

$$= \sqrt{E\left(\left[\sigma' \frac{1-2\tau'}{\tau'(1-\tau')} - (y_i - x_i'\beta') - \sigma \frac{1-2\tau}{\tau(1-\tau)} + (y_i - x_i'\beta) \right]^2 \right)}$$

$$= \sqrt{E\left(\left| \sigma' \frac{1-2\tau'}{\tau'(1-\tau')} - \sigma \frac{1-2\tau}{\tau(1-\tau)} + (x_i'(\beta - \beta')) \right|^2 \right)}$$

$$\leq \left(\left| \sigma' \frac{1-2\tau'}{\tau'(1-\tau')} - \sigma \frac{1-2\tau}{\tau(1-\tau)} \right|^2 \right)^{1/2} + (E|x_{i'}(\beta - \beta')|^2)^{1/2}$$

$$\leq \left(\left| \sigma' \frac{1-2\tau'}{\tau'(1-\tau')} - \sigma \frac{1-2\tau}{\tau(1-\tau)} \right|^2 \right)^{1/2} + ||\beta' - \beta|| (E||x_i||^2)^{1/2},$$

where the first inequality is given by Minkowski's inequality $(E|X+Y|^p)^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$ for $p \geq 1$, and the second inequality is Cauchy-Schwarz inequality. Hence, assumptions A3–A4 ensure that $\sup_{||\theta'-\theta||<\delta_n} \sum_{i=1}^{n} d_{2i} \to 0$ when $\delta_n \to 0$.

Finally, rewrite $\psi_{3\theta}(y, x) = (-\sigma + \rho_\tau(\gamma - x'\beta))$, and thus

$$d_{3i}(\theta', \theta) = \sqrt{E([\psi_{3\theta'} - \psi_{3\theta}]^2)}$$

$$= \sqrt{E([-\sigma' + \rho_{\tau'}(y_i - x_i\beta') + \sigma - \rho_\tau(y_i - x_i\beta)]^2)}$$

$$= \sqrt{E\left(\left[-\sigma' + \sigma + \rho_{\tau'}(y_i - x_i\beta') - \frac{1}{\sigma^2} \rho_\tau(y_i - x_i\beta) \right]^2 \right)}$$

$$\leq \sqrt{(-\sigma' + \sigma)^2} + \sqrt{E([\rho_{\tau'}(y_i - x_i\beta') - \rho_\tau(y_i - x_i\beta)]^2)}$$

$$= \sqrt{(-\sigma' + \sigma)^2} + \sqrt{E([\rho_{\tau'}(y_i - x_i\beta') - \rho_{\tau'}(y_i - x_i\beta) + \rho_{\tau'}(y_i - x_i\beta) - \rho_\tau(y_i - x_i\beta)]^2)}$$

$$\leq |\sigma - \sigma'| + \sqrt{E([||x_i(\beta' - \beta)|| + |\tau' - \tau|(y_i - x_i\beta)]^2)}$$

$$\begin{aligned}
&\leq|\sigma-\sigma'|+\sqrt{E([||x_i||||\beta'-\beta||+|\tau'-\tau|(y_i-x_i\beta)]^2)}\\
&\leq|\sigma-\sigma'|+(E[||x_i||||\beta'-\beta||]^2)^{1/2}+(E[|\tau'-\tau|(y_i-x_i\beta)]^2)^{1/2}\\
&=|\sigma-\sigma'|+||\beta'-\beta||(E||x_i||^2)^{1/2}+|\tau'-\tau|(E[(y_i-x_i\beta)]^2)^{1/2}\\
&\leq const\cdot(|\sigma-\sigma'|+||\beta'-\beta||+|\tau'-\tau|),
\end{aligned}$$

where the first inequality is given by Minkowski's inequality, the second inequality is given again by QR check function properties as $\rho_\tau(x+y)-\rho_\tau(y)\leq 2|x|$ and $\rho_{\tau_1}(y-x't)-\rho_{\tau_2}(y-x't)=(\tau_2-\tau_1)(y-x't)$. Third inequality is Cauchy-Schwarz inequality. Fourth is Minkowski's inequality. Last inequality uses assumption A4, and finally we have that $\sup_{||\theta'-\theta||<\delta_n}\sum_{i=1}^n d_{3i}\to 0$ when $\delta_n\to 0$.

Thus, $||\theta'-\theta||\to 0$ implies that $d(\theta',\theta)\to 0$ in every case, and therefore, $\sup_{||\theta'-\theta||<\delta_n}\sum_{i=1}^n d_i\to 0$ when $\delta_n\to 0$. The final condition in Theorem 2.11.1 in van der Vaart and Wellner (1996) is a Lindeberg condition, which is guaranteed by assumptions A1–A4. Therefore, we conclude that $\mathcal{F}$ is Donsker and

$$\sup_{||\theta-\theta_0||\leq\delta_n}||\mathbb{G}_n\psi_\theta(y,x)-\mathbb{G}_n\psi_{\theta_0}(y,x)||=o_p(1).$$ ■

References

Abadie, A., J. Angrist, and G. Imbens. 2002. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings." *Econometrica* 70: 91–117.

Angrist, J., V. Chernozhukov, and I. Fernández-Val. 2006. "Quantile Regression Under Misspecification, with an Application to the U.S. Wage Structure." *Econometrica* 74: 539–563.

Bloom, H. S. B., L. L. Orr, S. H. Bell, G. Cave, F. Doolittle, W. Lin, and J. M. Bos. 1997. "The Benefits and Costs of JTPA Title II-a Programs. Key Findings from the National Job Training Partnership Act Study." *Journal of Human Resources* 32: 549–576.

Chernozhukov, V., and C. Hansen. 2006. "Instrumental Quantile Regression Inference for Structural and Treatment Effects Models." *Journal of Econometrics* 132: 491–525.

Chernozhukov, V., and C. Hansen. 2008. "Instrumental Variable Quantile Regression: A Robust Inference Approach." *Journal of Econometrics* 142: 379–398.

Chernozhukov, V., I. Fernández-Val, and B. Melly. 2009. "Inference on Counterfactual Distributions." CEMMAP Working Paper CWP09/09.

Ebrahimi, N., E. S. Soofi, and R. Soyer. 2008. "Multivariate Maximum Entropy Identification, Transformation, and Dependence." *Journal of Multivariate Analysis* 99: 1217–1231.

Firpo, S. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica* 75: 259–276.

Geraci, M., and M. Botai. 2007. "Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution." *Biostatistics* 8: 140–154.

He, X., and Q.-M. Shao. 1996. "A General Bahadur Representation of M-Estimators and its Applications to Linear Regressions with Nonstochastic Designs." *Annals of Statistics* 24: 2608–2630.

He, X., and Q.-M. Shao. 2000. "Quantile Regression Estimates for a Class of Linear and Partially Linear Errors-in-Variables Models." *Statistica Sinica* 10: 129–140.

Hinkley, D. V., and N. S. Revankar. 1997. "Estimation of the Pareto Law from Underreported Data: A Further Analysis." *Journal of Econometrics* 5: 1–11.

Huber, P. J. 1967. "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions." In *Fifth Symposium on Mathematical Statistics and Probability*, 179–195. California: Unibersity of California, Berkeley.

Kitamura, Y., and M. Stutzer. 1997. "An Information-Theoretic Alternative to Generalized Method of Moments Estimation." *Econometrica* 65: 861–874.

Koenker, R. 2005. *Quantile Regression*. Cambridge: Cambridge University Press.

Koenker, R., and G. W. Bassett. 1978. "Regression Quantiles." *Econometrica* 46: 33–49.

Koenker, R., and J. A. F. Machado. 1999. "Godness of Fit and Related Inference Processes for Quantile Regression." *Journal of the American Statistical Association* 94: 1296–1310.

Koenker, R., and Z. Xiao. 2002. "Inference on the Quantile Regression Process." *Econometrica* 70: 1583–1612.

Komunjer, I. 2005. "Quasi-Maximum Likelihood Estimation for Conditional Quantiles." *Journal of Econometrics* 128: 137–164.

Komunjer, I. 2007. "Asymmetric Power Distribution: Theory and Applications to Risk Measurement." *Journal of Applied Econometrics* 22: 891–921.

Kosorok, M. R. 2008. *Introduction to Empirical Processes and Semiparametric Inference*. New York, New York: Springer-Verlag Press.

Kotz, S., T. J. Kozubowski, and K. Podgórsk. 2002a. "Maximum Entropy Characterization of Asymmetric Laplace Distribution." *International Mathematical Journal* 1: 31–35.

Kotz, S., T. J. Kozubowski, and K. Podgórsk. 2002b. "Maximum Likelihood Estimation of Asymmetric Laplace Distributions." *Annals of the Institute Statistical Mathematics* 54: 816–826.

LaLonde, R. J. 1995. "The Promise of Public-Sponsored Training Programs." *Journal of Economic Perspectives* 9: 149–168.

Machado, J. A. F. 1993. "Robust Model Selection and M-Estimation." *Econometric Theory* 9: 478–493.

Manski, C. F. 1991. "Regression." *Journal of Economic Literature* 29: 34–50.

Park, S. Y., and A. K. Bera. 2009. "Maximum Entropy Autoregressive Conditional Heteroskedasticity Model." *Journal of Econometrics* 150: 219–230.

Schennach, S. M. 2008. "Quantile Regression with Mismeasured Covariates." *Econometric Theory* 24: 1010–1043.

Soofi, E. S., and J. J. Retzer. 2002. "Information Indices: Unification and Applications." *Journal of Econometrics* 107: 17–40.

van der Vaart, A. 1998. *Asymptotic Statistics*. Cambridge: Cambridge University Press.

van der Vaart, A., and J. A. Wellner. 1996. *Weak Convergence and Empirical Processes*. New York, New York: Springer-Verlag.

Wei, Y., and R. J. Carroll. 2009. "Quantile Regression with Measurement Error." *Journal of the American Statistical Association* 104: 1129–1143.

Yu, K., and R. A. Moyeed. 2001. "Bayesian Quantile Regression." *Statistics & Probability Letters* 54: 437–447.

Yu, K., and J. Zhang. 2005. "A Three-Parameter Asymmetric Laplace Distribution and Its Extension." *Communications in Statistics – Theory and Methods* 34: 1867–1879.

Zhao, Z., and Z. Xiao. 2011. "Efficient Regressions Via Optimally Combining Quantile Information." Manuscript, University of Illinois at Urbana-Champaign.

Zou, H., and M. Yuan. 2008. "Composite Quantile Regression and the Oracle Model Selection Theory." *Annals of Statistics* 36: 1108–1126.